Neurocomputing 207 (2016) 300–309

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Kernel collaborative representation based dictionary learning and discriminative projection

Guoqing Zhang, Huaijiang Sun*, Guiyu Xia, Quansen Sun

School of Computer Science and Engineering, Nanjing University of Science and Technology, Jiangsu 210094, China

ARTICLE INFO

Article history:
Received 27 October 2015
Received in revised form
7 March 2016
Accepted 13 April 2016
Communicated by Huaping Liu
Available online 11 May 2016

Keywords:
Collaborative representation
Dictionary learning
Dimensionality reduction
Kernel method

ABSTRACT

Sparse representation based classification (SRC) has been developed and shown great potential due to its effectiveness in various classification tasks. But how to determine appropriate features that can best work with SRC remains an open question. Based on SRC and dimensionality reduction (DR) techniques, a simultaneous discriminative projection and dictionary learning method (DSRC) is proposed. However, as a linear algorithm, DSRC cannot handle the data with highly nonlinear distribution. Recently research has shown that the collaborative representation mechanism is more important to the success of SRC. Motivated by these concerns, in this paper, we propose a novel kernel collaborative representation based classifier (KCRC), and then we use it as a criterion to design a kernel collaborative representation based dictionary learning and discriminant projection method (KDL-DP). The proposed method aims at learning a projection matrix and a dictionary such that in the low dimension subspace the between-class reconstruction residual of a given data set is maximized and the within-class reconstruction residual is minimized. Extensive experimental results validate the superiority of the proposed approach when compared with the state-of-the-art methods.

1. Introduction

Dimensionality reduction (DR) is an important research topic in computer vision and pattern recognition fields. The high dimensional data usually lead to the inefficiency of many practical data processing techniques and may even degrade the performances of many classifiers. Many DR algorithms have been developed in the past decades [1–6]. The two most widely-used DR methods are principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2]. Both PCA and LDA are linear dimensionality reduction algorithms. They assume that the distributions of data sets are globally linear. However, high-dimensional data sets usually have non-linear structures in practical. Therefore, for conducting non-linear DR on a database, a family of manifold learning-related methods aroused wide research interests [3–6]. In addition, the kernel trick [7] has also been widely applied to extend linear DR algorithms to non-linear ones by performing linear operations on other higher or even infinite dimensional feature space induced by a kernel function. Yang et al. [8] proposed a complete kernel Fisher discriminant framework for feature extraction. Zafeiriou et al. [9] suggested a regularized kernel discriminant analysis method for face recognition and verification.

Recently, the idea of sparse representation is used to design some DR methods. Wright et al. [10] presented a sparse representation-based classification (SRC) method for face recognition. In SRC, a testing sample is sparsely reconstructed by the original training images and classified into the class with minimum sparse reconstruction residual. Qiao et al. [11] gave a sparsity preserving projections (SPP) for DR. In SPP, the graph (i.e. reconstruction relationships) was automatically constructed based on a modified sparse representation. Gui et al. [12] introduced the class information of data sets into SPP and presented a discriminant sparsity neighborhood preserving embedding (DSNPE). Yang et al. [13] claimed that DR algorithms should be designed according to the classifiers, so they gave a SRC steered discriminative projection (SRC-DP) method. By using SRC in the projected space, SRC-DP achieved better performance in face recognition. Clemmensen et al. [14] provided a sparse linear discriminant analysis (SLDA), which imposes sparseness constraint on projection vectors. The sparse projection vector yields a set of interpretable features for classification.

In sparse representation, the dictionary can be simply predefined as the training samples for discriminative sparse coding. However, it has been demonstrated that learning a dictionary from original training samples can provide better results [15–26]. These dictionary updating strategies are referred to as dictionary

* Corresponding author. Tel.: +86 13905172533.
E-mail addresses: xiayang14551@163.com (G. Zhang), sunhuaijiang@njust.edu.cn (H. Sun).

http://dx.doi.org/10.1016/j.neucom.2016.04.044
0925-2312/

learning (DL). In [16], a DL algorithm, K-SVD, is introduced that generalizes k-means clustering and efficiently learns an overcomplete dictionary from a set of training samples. Since K-SVD algorithm focuses on the represent power without considering the discrimination power. Thus, based on K-SVD, Zhang et al. [17] developed a discriminative K-SVD algorithm by simultaneously learning a linear classifier. Jiang et al. [18,19] introduced a label consistent regularization to enforce the discrimination of coding vectors. The so-called LC-KSVD algorithm exhibits good classification results. In order to further enhance the discrimination power of the learned dictionary, locality preserving K-SVD (LP-KSVD) [47] constructs a locality preserving term enforces the representation by local dictionary atoms similar to the input sample and penalizes the presentation by nonlocal (dissimilar) atoms. To exploit stronger discrimination, Yang et al. [21,48] proposed a Fisher discrimination DL method, where the category-specific strategy is adopted for learning a structural dictionary and the Fisher discrimination criterion is imposed on the coding vectors to enhance class discrimination. In addition, Remirez et al. [49] imposed a structure incoherence term on the dictionaries to make them independent to each other as much as possible. Cai et al. [50] proposed an SVM guided DL algorithm, which weights the squared distances between all pairs of coding vectors, where the weights are determined adaptively by training an SVM. Yang et al. [51] proposed a latent DL method by jointly learning a latent matrix to adaptively build the relationship between dictionary atoms and class labels.

Recently, several jointly DR and DL methods [22,52,53] have been proposed and reported more competitive performance than conventional DL methods. Lu et al. [53] proposed a learning framework which simultaneously learns the feature and the dictionary for image set based face recognition. Feng et al. [53] jointly learns a DR matrix and a discriminative dictionary for face recognition. Zhang et al. [22] presented a simultaneous projection and DL method using a carefully designed sigmoid reconstruction error (the ratio of intra-class error to inter-class error in sigmoid function).

Most of sparsity based classification schemes need to solve the ℓ_1-norm optimization problem, which is extremely time-consuming. Recent research in [27,28,45] have been shown that the collaborative representation (CR) mechanism is more important to the success of SRC. And a collaborative representation based classification (CRC) is given for image recognition by replacing the ℓ_1-norm regularization with the ℓ_2-norm regularization. CRC coded a testing sample by linear combination of all the training samples with regularized least square and classified the testing sample into the class with the minimum reconstruction error. Compared to SRC, CRC can achieve the competitive classification performance with much less computation burden.

All the sparse representation-based or collaborative representation-based methods mentioned above are designed based on a linear representation of the data. It is always inadequate for representing non-linear structures of the data. To deal with this problem, some non-linear methods have been proposed in literature [29–33]. These methods map the non-linear data into high-dimensional feature space using the kernel trick so that data of the same type are easily grouped together and are linearly separable. In this case we may find the sparse representation for the data more easily, and the reconstruction error may be reduced significantly as well [29–33]. Kernel sparse representation-based classifier (KSRC) is a non-linear extension of SRC and can deal with the problem occurred in SRC [29]. Lai et al. [34] proposed a kernel locality-constrained collaborative representation based discriminant analysis (KLCRC-DA), which achieved better results on face recognition. The well-known K-SVD [16] and method of optimal directions (MOD) [46] learning algorithms have also been adapted to the high dimensional feature space, and an efficient object recognition system that combines multiple classifiers based on the kernel sparse codes was presented in [33].

Motivated by the above concerns, in this paper, we propose a kernel collaborative representation based dictionary learning and discriminative projection method (KDL-DP). KDL-DP is designed according to the decision rule of our proposed kernel collaborative representation based classifier (KCRC), which is a nonlinear extension of CRC. The goal of the proposed method is to learn a projection matrix and a dictionary such that in the reduced subspace the within-class reconstruction residual is as small as possible and the between-class reconstruction residual is as large as possible. This encourages the sparse codes to be discriminative across classes. Therefore, KCRC can achieve better classification performance in the projected subspace.

The rest of paper is organized as follows: Section 2 reviews the CRC method. The kernel collaborative representation based classifier is proposed in Section 3. In Section 4, we design a kernel collaborative representation based dictionary learning and discriminative projection algorithm (KDL-DP). Experiments are described in Section 5 and Section 6 concludes the paper.

A note on notation: we use a bold upper case letter denotes a matrix, a bold lower case letter denotes a vector. For a matrix $\mathbf{P}$, $\mathbf{P}^T$ denotes its transpose. We summarize the key acronyms used throughout the paper in Table 1.

2. Collaborative representation based classifier

In this section, we briefly review collaborative representation based classification (CRC). Suppose we have c classes of subjects. Denote $\mathbf{X}_i = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{n_i}] \in R^{m\times n_i}$ as the training dataset of the $i-$th class, where $\mathbf{x}_j \in R^m, j = 1, \cdots, n_i$, is a $m-$ dimensional vector stretched by the $j-$th sample of class i. The entire training set is defined as $\mathbf{X} = [\mathbf{X}_1, \cdots, \mathbf{X}_c] \in R^{m\times n}$, where $n = \sum_{i=1}^{c} n_i$. Denote $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_c] \in R^{m\times q}$ as the structured dictionary, where $\mathbf{D}_i \in R^{m\times q_i}$ is the sub-dictionary associated with the $i-$th class $\mathbf{X}_i$, and $q = q_1 + \cdots + q_c$. Once a test sample $\mathbf{y}$ comes, we can represent $\mathbf{y}$ by using the dictionary, namely $\mathbf{y} = \mathbf{D}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1; \boldsymbol{\alpha}_2; \cdots; \boldsymbol{\alpha}_c]$ and $\boldsymbol{\alpha}_i$ is the reconstruction representation vector associated with class i. If $\mathbf{y}$ from the $i-$th class, usually $\mathbf{y} = \mathbf{D}_i\boldsymbol{\alpha}_i$ holds well. This implies that most coefficients in $\boldsymbol{\alpha}$ are nearly zeros and only $\boldsymbol{\alpha}_i$ has significant nonzero entries. Then the coding coefficient $\boldsymbol{\alpha}$ can be

Table 1
The key acronyms of our paper.

Notation	Meaning
DR	Dimensionality Reduction
DL	Dictionary Learning
SRC	Sparse Representation based Classification [10]
CRC	Collaborative representation based classification [27]
DSRC	Simultaneous discriminative projection and dictionary learning for sparse representation based classification [22]
SRC-DP	Sparse Representation Classifier Steered Discriminative Projection [13]
K-SVD	K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation [16]
LC-KSVD	Label Consistent K-SVD [19]
LP-KSVD	Joint kernel dictionary and classifier learning for sparse coding via locality preserving K-SVD [48]
FDDL	Fisher discrimination dictionary learning for sparse representation [21]
KSRC	Kernel Sparse Representation-based Classifier [29]
KCRC	Kernel collaborative representation based classifier
KDL-DP	Kernel collaborative representation based dictionary learning and discriminative projection